

Fuzzy Parameterized Queueing Models with Applications to Service Optimization

Dr. Nasim Akhtar

School Teacher(11-12), Department of Mathematics

UMV Jhauwan, Dighwara Saran, Bihar, India

nasimakhtar198304@gmail.com

Abstract: *Classical stochastic queueing models rely on precisely specified arrival and service parameters, an assumption that is often incompatible with service systems operating under incomplete information, measurement error, or linguistic assessments. In such environments, probabilistic descriptions alone may not adequately represent parameter uncertainty, leading to instability misclassification or suboptimal control decisions. This study formulates a queueing framework in which key system parameters, including arrival rates and service capacities, are represented as fuzzy numbers within a stochastic setting. The model integrates α -cut representations with standard queueing performance measures, enabling the derivation of stability conditions under imprecise parameterization. Sufficient conditions for system stability are established by examining fuzzy traffic intensity and its admissible realizations. Building on these results, an optimization problem is posed to minimize expected waiting-related costs subject to fuzzy stability and resource constraints. The formulation admits tractable solution procedures based on parametric decomposition across α -levels. A numerical illustration motivated by a hospital outpatient department demonstrates how fuzzy parameterization captures operational ambiguity in demand intensity and service efficiency, and how optimized staffing decisions differ from those obtained under crisp assumptions. The analysis shows that incorporating fuzziness leads to more robust performance assessments while preserving analytical transparency*

Keywords: Fuzzy queueing systems; stability analysis; optimization under uncertainty; service systems; waiting time

I. INTRODUCTION

Service systems such as banks, hospitals, call centers, and transportation hubs operate in environments where demand and service capabilities exhibit significant variability and partial observability. Queueing theory has long provided a mathematical foundation for analyzing congestion, delay, and resource utilization in such systems. Classical models based on stochastic processes describe arrivals and services through precisely specified probability distributions, yielding analytically tractable performance measures such as waiting times, queue lengths, and system stability criteria [1, 2]. These models have supported decision-making in a wide range of operational contexts, from capacity planning to scheduling and admission control.

Despite their mathematical elegance, classical stochastic queueing models rest on an assumption that key parameters, such as arrival rates or service intensities, are known accurately. In many real-world service environments, this assumption is difficult to justify. Empirical data may be scarce, non-stationary, or contaminated by measurement error, while expert assessments of demand or service efficiency are often expressed in qualitative or linguistic terms rather than precise numerical values. As a consequence, crisp parameter specifications may lead to misleading stability assessments or overly confident optimization outcomes, particularly when systems operate close to critical loading.

Probabilistic modeling captures randomness arising from intrinsic variability, but it does not address epistemic uncertainty associated with incomplete knowledge of system parameters. In response to this limitation, several researchers have sought to enrich queueing analysis by incorporating alternative representations of uncertainty. Fuzzy



set theory, introduced to formalize imprecision and vagueness in mathematical models, provides a natural framework for representing uncertain parameters through membership functions rather than point estimates [7]. Within this framework, arrival rates or service capacities can be expressed as fuzzy numbers, reflecting ranges of plausible values together with degrees of confidence. Such representations are particularly relevant in service systems where parameters are estimated from limited historical data or elicited from human judgment.

The integration of fuzzy concepts into queueing theory has attracted sustained attention over the past decades. Early contributions focused on replacing crisp parameters in classical models with fuzzy counterparts and evaluating performance measures through α -cut techniques or possibility-based reasoning [8, 9]. Subsequent studies examined fuzzy arrival and service processes, fuzzy waiting times, and decision-making rules under fuzzy congestion measures [11, 10]. These works demonstrated that fuzzy parameterization can accommodate imprecision without abandoning the structural insights provided by stochastic queueing models. At the same time, they highlighted methodological challenges related to stability characterization and optimization when system parameters are no longer single-valued.

Stability analysis plays a central role in queueing theory, as it delineates the operational regimes under which performance measures remain finite. In classical settings, stability conditions are typically expressed in terms of traffic intensity ratios derived from exact arrival and service rates [3]. When these parameters are fuzzy, the notion of stability becomes inherently set-valued, raising questions about admissible operating regions and the interpretation of borderline cases. Similarly, optimization problems in service systems often seek to minimize waiting times or operational costs subject to stability constraints. Under fuzzy parameterization, such problems require reformulation to ensure feasibility across admissible parameter realizations while retaining computational tractability.

The present work addresses these issues by developing a fuzzy-stochastic queueing framework that preserves the analytical structure of classical models while explicitly accounting for parameter imprecision. Fuzzy numbers are employed to represent uncertain arrival and service parameters, and α -cut representations are used to derive stability conditions that are consistent with standard queueing theory at each confidence level. Building on these results, an optimization formulation is constructed to determine service capacities that minimize expected waiting-related costs under fuzzy stability constraints. The approach remains grounded in established queueing principles, avoiding reliance on ad hoc performance indices.

The relevance of the proposed framework is illustrated through an application to a healthcare service system, where outpatient demand and service efficiency are subject to significant uncertainty due to patient heterogeneity and operational variability. By comparing fuzzy-based decisions with those obtained under crisp assumptions, the analysis highlights the impact of parameter imprecision on stability assessment and resource allocation.

The remainder of the paper is organized as follows. Section 2 introduces the mathematical preliminaries on queueing models and fuzzy numbers required for the subsequent analysis. Section 3 formulates the fuzzy parameterized queueing model and derives stability conditions under uncertainty. Section 4 presents an optimization framework for service capacity decisions and discusses its solution via α -level decomposition. Section 5 provides a numerical illustration motivated by a hospital service setting. Section 6 concludes with a discussion of implications and directions for further research.

II. MATHEMATICAL PRELIMINARIES

This section summarizes the mathematical concepts from queueing theory and fuzzy set theory required for the development of the proposed model. The presentation is restricted to standard definitions and notation that will be used consistently throughout the paper.

2.1 Basic Queueing Notation

A queueing system is typically characterized by an arrival process, a service mechanism, and a queue discipline. Let $\{A(t), t \geq 0\}$ denote the counting process representing the number of arrivals up to time t , and let $\{S_n\}_{n \geq 1}$ be a sequence



of independent service times. In classical $M/M/1$ or $M/M/c$ models, interarrival and service times are assumed to follow exponential distributions with parameters $\lambda > 0$ and $\mu > 0$, respectively [1, 2].

The traffic intensity is defined as

$$\rho = \frac{\lambda}{c\mu}, \quad (1)$$

where c denotes the number of parallel servers. For a single-server system ($c = 1$), stability requires $\rho < 1$, ensuring the existence of a stationary distribution for the queue length process. Under stability, standard performance measures include the expected queue length L_q and the expected waiting time W_q , related by Little's law,

$$L_q = \lambda W_q. \quad (2)$$

2.2 Fuzzy Sets and Membership Functions

Let X be a universal set. A fuzzy set \tilde{A} in X is defined by a membership function $\mu_{\tilde{A}} : X \rightarrow [0, 1]$, where $\mu_{\tilde{A}}(x)$ represents the degree to which the element $x \in X$ belongs to \tilde{A} [7]. Formally,

$$\tilde{A} = \{x, \mu_{\tilde{A}}(x) : x \in X\}. \quad (3)$$

In contrast to crisp sets, fuzzy sets allow partial membership, which provides a mathematical representation of imprecision or vagueness. Basic operations such as union and intersection are defined pointwise using appropriate t-norms and t-conorms. For instance, the intersection of two fuzzy sets \tilde{A} and \tilde{B} is often defined as

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = \min\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\}. \quad (4)$$

2.3 Fuzzy Numbers and α -Cuts

A fuzzy number is a special class of fuzzy sets defined on the real line that satisfies normality, convexity, and upper semi-continuity [8]. A fuzzy number \tilde{x} is commonly represented by its membership function $\mu_{\tilde{x}} : \mathbb{R} \rightarrow [0, 1]$, with bounded support.

An important analytical tool is the α -cut of a fuzzy number, defined for $\alpha \in (0, 1]$ as

$$[\tilde{x}]_{\alpha} = \{x \in \mathbb{R} : \mu_{\tilde{x}}(x) \geq \alpha\} = [x_L(\alpha), x_U(\alpha)]. \quad (5)$$

The α -cut yields a closed interval representing all values of x compatible with the confidence level α . Arithmetic operations on fuzzy numbers can be performed through interval arithmetic applied to their α -cuts, allowing fuzzy quantities to be propagated through queueing performance expressions in a systematic manner [9].

2.4 Defuzzification Concepts

Defuzzification refers to the process of mapping a fuzzy number to a representative crisp value, typically for decision-making purposes. One commonly used approach is the centroid (or center of gravity) method, defined as

$$x^* = \frac{\int_{\mathbb{R}} x \mu_{\tilde{x}}(x) dx}{\int_{\mathbb{R}} \mu_{\tilde{x}}(x) dx}. \quad (6)$$

Alternative methods include mean of maxima or α -level averaging [10]. In the present study, defuzzification is employed only at the final decision stage, while stability analysis and optimization are carried out primarily in terms of α -cuts to preserve the underlying uncertainty structure.

III. FUZZY PARAMETERIZED QUEUEING MODEL

This section introduces a queueing model in which uncertainty in system parameters is represented through fuzzy numbers. The construction preserves the probabilistic structure of classical queueing systems while allowing for imprecision in arrival and service characteristics.



3.1 Classical Queueing Framework

Consider a standard $M/M/c$ queue with infinite buffer and first-come, first-served discipline. Customers arrive according to a Poisson process with rate $\lambda > 0$, and service times are independent and exponentially distributed with rate $\mu > 0$ at each of the c identical servers. Let

$N(t)$ denote the number of customers in the system at time t . Under the classical formulation, $\{N(t), t \geq 0\}$ forms a continuous-time Markov chain with birth rates λ and death rates $\min\{n, c\}\mu$.

The classical traffic intensity is defined as

$$\rho = \frac{\lambda}{c\mu}, \quad (7)$$

and the system is stable if $\rho < 1$. Under this condition, steady-state performance measures such as the expected waiting time W_q and expected system size L are finite and admit closed-form expressions [2, 3].

3.2 Fuzzy Arrival and Service Rates

In many service systems, the precise values of λ and μ are not known with certainty. Instead, they are often estimated from limited data or expressed linguistically (e.g., “moderate arrival intensity” or “high service efficiency”). To model this imprecision, the arrival rate and service rate are represented as fuzzy numbers.

Definition 3.1. The fuzzy arrival rate is denoted by $\tilde{\lambda}$, and the fuzzy service rate by $\tilde{\mu}$. Both are assumed to be fuzzy numbers defined on \mathbb{R}^+ with bounded support.

Under this representation, the queueing system retains its stochastic dynamics conditional on particular realizations of (λ, μ) , while uncertainty is encoded through the membership functions of $\tilde{\lambda}$ and $\tilde{\mu}$.

3.3 Triangular Fuzzy Representation

For analytical transparency, triangular fuzzy numbers are adopted to describe uncertain parameters. A triangular fuzzy number $\tilde{x} = (x_1, x_2, x_3)$ is defined by the membership function

$$\mu_{\tilde{x}}(x) = \begin{cases} 0, & x < x_1 \\ \frac{x - x_1}{x_2 - x_1}, & x_1 \leq x \\ \frac{x_3 - x}{x_3 - x_2}, & x_2 \leq x \\ 0, & x > x_3 \end{cases}, \quad (8)$$

where $x_1 < x_2 < x_3$ represent the lower bound, most plausible value, and upper bound, respectively.

Accordingly, the fuzzy arrival and service rates are specified as

$$\tilde{\lambda} = (\lambda_1, \lambda_2, \lambda_3), \quad \tilde{\mu} = (\mu_1, \mu_2, \mu_3), \quad (9)$$

reflecting imprecision in demand intensity and service capacity.

3.4 α -Cut Reformulation

To facilitate analysis, fuzzy parameters are handled through their α -cuts. For $\alpha \in (0, 1]$, the α -cuts of $\tilde{\lambda}$ and $\tilde{\mu}$ are given by

$$[\tilde{\lambda}]_{\alpha} = [\lambda_L(\alpha), \lambda_U(\alpha)], \quad [\tilde{\mu}]_{\alpha} = [\mu_L(\alpha), \mu_U(\alpha)], \quad (10)$$

where

$$\lambda_L(\alpha) = \lambda_1 + \alpha(\lambda_2 - \lambda_1), \quad \lambda_U(\alpha) = \lambda_3 - \alpha(\lambda_3 - \lambda_2), \quad (11)$$

and analogous expressions hold for $\mu_L(\alpha)$ and $\mu_U(\alpha)$.



At each α -level, the fuzzy queue reduces to a family of classical $M/M/c$ systems parameterized by $(\lambda, \mu) \in [\tilde{\lambda}]_{\alpha} \times [\tilde{\mu}]_{\alpha}$. Performance measures and stability conditions can therefore be examined through interval analysis.

3.5 Assumptions

The following assumptions are imposed throughout the analysis:

- (A1) Conditional on fixed (λ, μ) , arrivals follow a Poisson process and service times are exponentially distributed.
- (A2) The fuzzy parameters $\tilde{\lambda}$ and $\tilde{\mu}$ are independent fuzzy numbers.
- (A3) The number of servers c is deterministic and known.
- (A4) Fuzziness captures epistemic uncertainty in parameter specification and does not replace the stochastic nature of arrivals or services.

3.6 Remarks

Remark 3.2. The α -cut formulation ensures consistency with classical queueing theory: for $\alpha = 1$, the model reduces to a crisp $M/M/c$ system with parameters (λ_2, μ_2) .

Remark 3.3. By separating stochastic variability from parameter imprecision, the proposed framework allows stability and optimization analyses to be conducted using established queueing results while explicitly accounting for uncertainty.

IV. STABILITY ANALYSIS OF THE FUZZY QUEUEING SYSTEM

Stability is a fundamental requirement in queueing analysis, as it determines whether steady-state performance measures exist. This section extends classical stability conditions to queueing systems with fuzzy-parameterized arrival and service rates.

4.1 Classical Stability Condition

For a classical $M/M/c$ queue with deterministic parameters (λ, μ) , stability is guaranteed if and only if the traffic intensity satisfies

$$\rho = \frac{\lambda}{c\mu} < 1. \quad (12)$$

Under this condition, the underlying birth-death process is positive recurrent, and the stationary distribution of the queue length exists [3]. When $\rho \geq 1$, the expected queue length and waiting time diverge.

4.2 Fuzzy Stability Concept

When the arrival and service rates are represented as fuzzy numbers $\tilde{\lambda}$ and $\tilde{\mu}$, the traffic intensity becomes a fuzzy quantity,

$$\tilde{\rho} = \frac{\tilde{\lambda}}{c\tilde{\mu}}. \quad (13)$$

In this setting, stability cannot be assessed using a single scalar inequality. Instead, stability must be interpreted in terms of admissible realizations of (λ, μ) contained within the fuzzy parameter space. Intuitively, the system is regarded as stable at a given confidence level if all plausible parameter realizations at that level satisfy the classical stability condition.

4.3 Stability under α -Cuts

Let $\alpha \in (0, 1]$ and consider the α -cuts

$$[\tilde{\lambda}]_{\alpha} = [\lambda_L(\alpha), \lambda_U(\alpha)], \quad [\tilde{\mu}]_{\alpha} = [\mu_L(\alpha), \mu_U(\alpha)].$$

For each α , the fuzzy queue induces a family of crisp $M/M/c$ systems indexed by (λ, μ) in the rectangle

$$P_{\alpha} = [\lambda_L(\alpha), \lambda_U(\alpha)] \times [\mu_L(\alpha), \mu_U(\alpha)].$$



Define the α -level traffic intensity interval as

$$[\rho]_{\alpha} = \left[\frac{\lambda_L(\alpha)}{c\mu_U(\alpha)}, \frac{\lambda_U(\alpha)}{c\mu_L(\alpha)} \right]. \quad (14)$$

Stability at level α requires that the upper bound of this interval be strictly less than one.

Theorem 4.1 (Fuzzy Stability Condition). For a fuzzy $M/M/c$ queue with fuzzy arrival rate $\tilde{\lambda}$ and fuzzy service rate $\tilde{\mu}$, the system is stable at confidence level $\alpha \in (0, 1]$ if

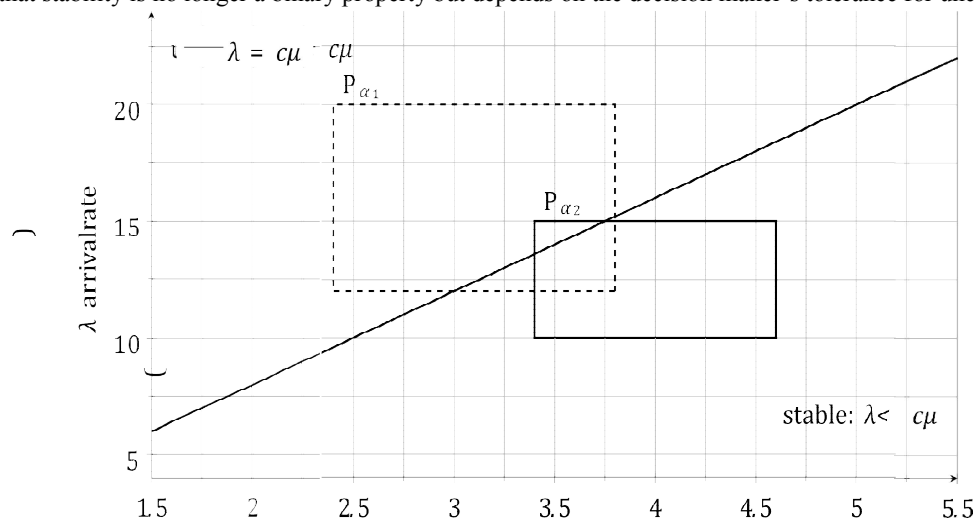
$$\frac{\lambda_U(\alpha)}{c\mu_L(\alpha)} < 1. \quad (15)$$

If this inequality fails for some α , then instability cannot be excluded at that confidence level.

Proof Sketch. At a fixed α -level, the fuzzy queue reduces to a family of classical $M/M/c$ systems with parameters in P_{α} . For any realization $(\lambda, \mu) \in P_{\alpha}$, stability requires $\lambda/(c\mu) < 1$. The worst-case traffic intensity over P_{α} is attained at $(\lambda_U(\alpha), \mu_L(\alpha))$, yielding the stated condition. If this worst-case system is stable, then all admissible realizations at level α are stable. Conversely, violation of the condition implies the existence of at least one unstable realization. \square

4.4 Interpretation of the Stability Region

The fuzzy stability condition defines a confidence-dependent feasible region in the (λ, μ) plane. For larger values of α , the parameter intervals shrink toward their most plausible values, resulting in a less conservative stability requirement. For smaller α , wider parameter ranges lead to stricter conditions, reflecting increased uncertainty. This interpretation highlights that stability is no longer a binary property but depends on the decision maker's tolerance for uncertainty.



μ (service rate per server)

Figure 1: Illustration of the stability region in the (λ, μ) plane. The curve $\lambda = c\mu$ separates feasible (stable) and infeasible (unstable) regions. Rectangular α -cut parameter sets intersecting the unstable region indicate loss of stability at the corresponding confidence level.

V. FUZZY PERFORMANCE MEASURES

This section develops fuzzy counterparts of standard performance measures for an $M/M/c$ queue with fuzzy arrival and service rates. The construction is based on α -cuts, which reduce the problem to interval evaluation of classical expressions at each confidence level.



5.1 Classical Performance Measures for $M/M/c$

Fix crisp parameters (λ, μ) with $\rho = \lambda/(c\mu) < 1$. Let

$$a = \frac{\lambda}{\mu}, \quad \rho = \frac{a}{c}. \quad (16)$$

Define the normalizing constant

$$P_0(\lambda, \mu; c) = \left(\sum_{n=0}^{c-1} \frac{a^n}{n!} + \frac{a^c}{c!} \cdot \frac{1}{1-\rho} \right)^{-1}, \quad (17)$$

and the Erlang-C delay probability

$$P_{\text{wait}}(\lambda, \mu; c) = \frac{a^c}{c!} \cdot \frac{1}{1-\rho} P_0. \quad (18)$$

Then the expected queue length and waiting time are [2, 1]

$$L_q(\lambda, \mu; c) = \frac{P_{\text{wait}}(\lambda, \mu; c) \rho}{1-\rho}, \quad (19)$$

$c\mu^{-1}; c$

$$W_q(\lambda, \mu; c) = \frac{L_q(\lambda, \mu; c)}{\lambda} = \frac{P_{\text{wait}}(\lambda, \mu; c) \rho}{\lambda(1-\rho)}. \quad (20)$$

The expected system time and utilization are

$$W(\lambda, \mu; c) = W_q(\lambda, \mu; c) + \frac{1}{\mu}, \quad (21)$$

$$U(\lambda, \mu; c) = \rho = \frac{\lambda}{c\mu}. \quad (22)$$

Finally, Little's law yields the expected number in system,

$$L(\lambda, \mu; c) = \lambda W(\lambda, \mu; c) = L. \quad (23)$$

5.2 α -Cut Performance Measures under Fuzzy Parameters

Let $\tilde{\lambda}$ and $\tilde{\mu}$ be fuzzy numbers on \mathbb{R}^+ with α -cuts

$$[\tilde{\lambda}]_\alpha = [\lambda_L(\alpha), \lambda_U(\alpha)], \quad [\tilde{\mu}]_\alpha = [\mu_L(\alpha), \mu_U(\alpha)], \quad \alpha \in (0, 1]. \quad (24)$$

For a fixed α , define the admissible parameter set

$$P_\alpha = [\lambda_L(\alpha), \lambda_U(\alpha)] \times [\mu_L(\alpha), \mu_U(\alpha)]. \quad (25)$$

To ensure finiteness of queueing measures at level α , impose the α -level stability condition

$$\frac{\lambda_U(\alpha)}{c \mu_L(\alpha)} < 1, \quad (26)$$

as established in Section 4.

For any performance functional $g(\lambda, \mu)$ defined on the stable region, the α -cut of the induced fuzzy quantity \tilde{g} is defined by

$$[\tilde{g}]_\alpha = [g_L(\alpha), g_U(\alpha)] = \left[\min_{(\lambda, \mu) \in P_\alpha} g(\lambda, \mu), \max_{(\lambda, \mu) \in P_\alpha} g(\lambda, \mu) \right]. \quad (27)$$

Because the exact minimizers and maximizers can be nontrivial for $M/M/c$ expressions (through (17)–(18)), the analysis below identifies extremal bounds that are consistent with monotonicity properties and are suitable for interval-based computation.



5.3 Fuzzy Utilization

Utilization has the simplest structure. Since $U(\lambda, \mu; c) = \lambda/(c\mu)$ is increasing in λ and decreasing in μ on \mathbb{R}_+^2 , its α -cut is

$$[\tilde{U}]_\alpha = \left[\frac{\lambda_L(\alpha)}{c\mu_U(\alpha)}, \frac{\lambda_U(\alpha)}{c\mu_L(\alpha)} \right], \quad (28)$$

with the upper endpoint strictly less than one under (26).

5.4 Fuzzy Waiting Time and Queue Length

The waiting time $W_q(\lambda, \mu; c)$ in (20) is increasing in λ and decreasing in μ within the stable region, reflecting that higher demand and slower service increase delay. Consequently, conservative α -cut bounds can be obtained by extremizing at the corners of P_α :

$$[\tilde{W}_q]_\alpha \subseteq [W_q(\lambda_L(\alpha), \mu_U(\alpha); c), W_q(\lambda_U(\alpha), \mu_L(\alpha); c)], \quad (29)$$

where the right endpoint is finite if (26) holds. In computations, one may evaluate W_q at these corners to produce practical bounds; the inclusion becomes equality under mild regularity conditions when monotonicity is strict across P_α . Similarly, $L_q(\lambda, \mu; c)$ in (19) is increasing in λ and decreasing in μ in the stable region, yielding

$$[\tilde{L}_q]_\alpha \subseteq [L_q(\lambda_L(\alpha), \mu_U(\alpha); c), L_q(\lambda_U(\alpha), \mu_L(\alpha); c)]. \quad (30)$$

An alternative is to compute $[L_q]_\alpha$ from $[W_q]_\alpha$ using interval multiplication with $[\tilde{\lambda}]_\alpha$, via Little's law, but direct evaluation through (19) avoids compounding interval conservatism.

5.5 Fuzzy System Time and Number in System

The expected system time is the sum of waiting time and mean service time,

$$W(\lambda, \mu; c) = W_q(\lambda, \mu; c) + \frac{1}{\mu}. \quad (31)$$

Using α -cut arithmetic, a conservative enclosure is

$$[\tilde{W}]_\alpha \subseteq [\tilde{W}_q]_\alpha \oplus \left[\frac{1}{\mu_U(\alpha)}, \frac{1}{\mu_L(\alpha)} \right], \quad (32)$$

where \oplus denotes interval addition. Since $1/\mu$ is decreasing in μ , its extremal values occur at $\mu_U(\alpha)$ and $\mu_L(\alpha)$.

For the expected number in system, $L(\lambda, \mu; c) = \lambda W(\lambda, \mu; c)$, interval multiplication gives

$$[\tilde{L}]_\alpha \subseteq [\tilde{\lambda}]_\alpha \otimes [\tilde{W}]_\alpha \quad (33)$$

where \otimes denotes interval multiplication. Because all intervals here are nonnegative, this simplifies to

$$[\tilde{L}]_\alpha \subseteq [\lambda_L(\alpha)W_L(\alpha), \lambda_U(\alpha)W_U(\alpha)], \quad (34)$$

with $W_L(\alpha)$ and $W_U(\alpha)$ denoting the endpoints of $[\tilde{W}]_\alpha$.

5.6 Summary Table of α -Level Formulas

Table 1 summarizes the α -cut enclosures used for the fuzzy performance measures. Each entry is computed under the α -level stability condition (26).

Table 1: Summary of α -cut enclosures for fuzzy performance measures in an $M/M/c$ system with fuzzy parameters.

| Measure | α -cut enclosure |
|--------------------|--|
| Utilization U | $[\tilde{U}]_\alpha = \left[\frac{\lambda_L(\alpha)}{c\mu_U(\alpha)}, \frac{\lambda_U(\alpha)}{c\mu_L(\alpha)} \right]$ |
| Waiting time W_q | $[\tilde{W}_q]_\alpha \subseteq [W_q(\lambda_L(\alpha), \mu_U(\alpha); c), W_q(\lambda_U(\alpha), \mu_L(\alpha); c)]$ |
| Queue length L_q | $[\tilde{L}_q]_\alpha \subseteq [L_q(\lambda_L(\alpha), \mu_U(\alpha); c), L_q(\lambda_U(\alpha), \mu_L(\alpha); c)]$ |
| System time W | $[\tilde{W}]_\alpha \subseteq [\frac{1}{\mu_U(\alpha)}, \frac{1}{\mu_L(\alpha)}]$ |
| Number in system | $[\tilde{L}]_\alpha \subseteq [\lambda_L(\alpha)W_L(\alpha), \lambda_U(\alpha)W_U(\alpha)]$ |



$$[\tilde{W}^*]_\alpha \subseteq [\tilde{W}_q]_\alpha \oplus \left[\frac{1}{\mu_U(\alpha)} \right],$$

$$\tilde{L} \quad [\tilde{L}]_\alpha \subseteq [\tilde{\lambda}]_\alpha \otimes [\tilde{W}]_\alpha$$

Remark 5.1. The enclosures in Table 1 are designed for transparent computation and interpretation. In implementations, one may replace the inclusions by equalities by solving the supremum/infimum problems in (27) numerically over P_α , which can be done efficiently since P_α is a compact rectangle and the classical formulas are smooth on the stable region.

VI. SERVICE OPTIMIZATION UNDER FUZZINESS

This section formulates a service optimization problem for a fuzzy-parameterized $M/M/c$ system. The decision aims to allocate service capacity while accounting for epistemic uncertainty in arrival and service rates. The emphasis is on an interpretable model whose feasibility is governed by the α -level stability condition and whose objective reflects waiting-related operational costs.

6.1 Decision Variables and Control Setting

Let $c \in \mathbb{N}$ denote the number of parallel servers. The design problem considered here is a capacity decision: select c to balance congestion and staffing costs. In applications, c may represent the number of tellers in a bank branch, physicians in an outpatient clinic, or agents in a call center. The uncertain arrival and service rates are modeled as fuzzy numbers $\tilde{\lambda}$ and $\tilde{\mu}$ with α -cuts

$$[\tilde{\lambda}]_\alpha = [\lambda_L(\alpha), \lambda_U(\alpha)], \quad [\tilde{\mu}]_\alpha = [\mu_L(\alpha), \mu_U(\alpha)], \quad \alpha \in (0, 1],$$

as in Section 3. For each feasible c , the classical waiting time is given by the $M/M/c$ formula $W_q(\lambda, \mu; c)$ in (20).

6.2 Fuzzy Objective: Waiting–Cost Tradeoff

A common operational objective is to reduce congestion while controlling staffing costs. Let $h > 0$ denote the waiting cost rate per customer per unit time (measured in monetary units), and let $s > 0$ denote the cost per server per unit time. For crisp parameters, a standard steady–state cost proxy is

$$J(\lambda, \mu; c) = hL_q(\lambda, \mu; c) + sc, \quad (35)$$

where $L_q(\lambda, \mu; c)$ is given by (19). Under fuzzy parameters, $J(\lambda, \mu; c)$ induces a fuzzy cost $\tilde{J}(c)$ whose α -cut can be defined via

$$[\tilde{J}(c)]_\alpha = \left[\inf_{(\lambda, \mu) \in P_\alpha} (hL_q(\lambda, \mu; c) + sc), \right], \quad (36)$$

where $P_\alpha = [\tilde{\lambda}]_\alpha \times [\tilde{\mu}]_\alpha$.

In planning settings, it is often desirable to control the worst plausible congestion cost at each confidence level. This motivates the α -level robust objective

$$J^U(\alpha; c) = \sup_{(\lambda, \mu) \in P_\alpha} (hL_q(\lambda, \mu; c) + sc), \quad (37)$$

with

$$L_q^U(\alpha; c) = \sup_{(\lambda, \mu) \in P_\alpha} L_q(\lambda, \mu; c). \quad (38)$$

A corresponding conservative lower bound $J^L(\alpha; c)$ can be defined using the infimum in (36), but it is the upper bound that typically governs risk-averse capacity planning.

6.3 Constraints: Stability and Capacity Limits

Feasibility requires stability across the admissible parameter set at the chosen confidence level.

Using the α -level stability condition from Section 4, we enforce



$$\frac{\lambda_U(\alpha)}{c \mu_L(\alpha)} \leq 1 - \varepsilon, \quad (39)$$

for a chosen margin $\varepsilon \in (0,1)$ that prevents near-critical loading (which can lead to large delays even when stable). In addition, practical considerations impose bounds

$$c_{\min} \leq c \leq c_{\max}, \quad c \in \mathbb{N}, \quad (40)$$

where c_{\min} and c_{\max} reflect minimum staffing requirements and resource availability.

6.4 Optimization Model

Fix a planning confidence level $\alpha \in (0,1]$. The fuzzy capacity design problem is formulated as the discrete optimization problem

$$\min_{c \in \mathbb{N}} J^U(\alpha; c) = h L_q^U(\alpha; c) + s c \quad (41) \text{ s.t. } \frac{\lambda_U(\alpha)}{c \mu_L(\alpha)} \leq 1 - \varepsilon, \quad (42)$$

$$c_{\min} \leq c \leq c_{\max}. \quad (43)$$

The model selects the number of servers that minimizes a worst-plausible steady-state congestion cost plus staffing cost, subject to stability over the α -cut parameter set.

Remark 6.1. Model (41)–(43) may be written in waiting-time form by replacing $L_q(\lambda, \mu; c)$ with $\lambda W_q(\lambda, \mu; c)$ via Little's law, resulting in a fuzzy objective based on W_q . The cost form (35) is convenient for service applications where staffing costs are explicit.

6.5 Solution Strategy via α -Level Decomposition

Two structural features simplify the solution of (41)–(43). First, the feasibility constraint (42) yields a closed-form lower bound on admissible c :

$$c \geq \left\lceil \frac{\lambda_U(\alpha)}{(1 - \varepsilon) \mu_L(\alpha)} \right\rceil. \quad (44)$$

Hence the feasible set is a finite integer interval once c_{\max} is specified. Second, for fixed c , the upper bound $L_q^U(\alpha; c)$ in (38) can be computed by evaluating $L_q(\lambda, \mu; c)$ over the compact rectangle P_α .

In many $M/M/c$ settings, $L_q(\lambda, \mu; c)$ is increasing in λ and decreasing in μ on the stable region, suggesting the conservative evaluation

$$L_q^U(\alpha; c) \approx L_q(\lambda_U(\alpha), \mu_L(\alpha)), \quad (45)$$

which is consistent with the α -cut enclosures in Section 5. When higher accuracy is required,

$L_q^U(\alpha; c)$ can be obtained by a two-dimensional search over P_α since the domain is rectangular and the classical formulas are smooth away from the boundary $\lambda = c\mu$. In either case, the outer optimization over c can be solved by enumeration over feasible integers:

$$c^*(\alpha) \in \arg \min_{c \in C_\alpha} \{h L_q^U(\alpha; c)\}, \quad C_\alpha = \{c \in \mathbb{N} : (42), (43) \text{ hold}\}. \quad (46)$$

Remark 6.2. The dependence of $c^*(\alpha)$ on α provides a transparent sensitivity analysis with respect to parameter imprecision. Smaller α corresponds to wider parameter ranges and typically leads to larger feasible lower bounds (44), reflecting more conservative staffing to maintain stability and control delay.

VII. NUMERICAL ILLUSTRATION

This section illustrates the proposed fuzzy queueing and optimization framework using a stylized hospital outpatient service unit. The setting reflects operational conditions where demand intensity and service efficiency are estimated from limited observations and expert judgment.



7.1 Service Setting and Parameter Specification

Consider a walk-in outpatient department operating with c parallel physicians during a fixed shift. Patient arrivals exhibit day-to-day variability and are influenced by seasonal factors; service times vary with case mix. To represent imprecision in parameter estimation, the arrival and service rates are modeled as triangular fuzzy numbers:

$$\begin{aligned} \hat{\lambda} &= (\lambda_1, \lambda_2, \lambda_3) = (9.0, 10.5, 12.0) \text{ patients/hour, (47)} \\ \tilde{\mu} &= (\mu_1, \mu_2, \mu_3) = (3.2, 3.6, 4.0) \text{ patients/hour/server. (48)} \end{aligned}$$

The modal values (λ_2, μ_2) correspond to typical operating conditions inferred from historical observations, whereas (λ_1, λ_3) and (μ_1, μ_3) reflect plausible deviations due to demand surges and service slowdowns or speedups.

For each $\alpha \in (0, 1]$, the α -cuts are

$$[\hat{\lambda}]_\alpha = [\lambda_L(\alpha), \lambda_U(\alpha)] = [9.0 + 1.5\alpha, 12.0 - 1.5\alpha], \quad (49)$$

$$[\tilde{\mu}]_\alpha = [\mu_L(\alpha), \mu_U(\alpha)] = [3.2 + 0.4\alpha, 4.0 - 0.4\alpha]. \quad (50)$$

We examine $\alpha \in \{0.2, 0.5, 0.8, 1.0\}$ as representative confidence levels. At each α , the stability requirement for an $M/M/c$ system is imposed using the worst-case inequality

$$\frac{\lambda_U(\alpha)}{c \mu_L(\alpha)} < 1 \quad (51)$$

Given stability, the waiting time $W_q(\lambda, \mu; c)$ is computed using the classical Erlang-C expression (20). To obtain interpretable fuzzy bounds, we evaluate the corner systems

$$W_q^L(\alpha; c) = W_q(\lambda_L(\alpha), \mu_U(\alpha); c), \quad W_q^U(\alpha; c) = W_q(\lambda_U(\alpha), \mu_L(\alpha); c), \quad (52)$$

which provide an α -level enclosure consistent with Section 5.

7.2 Computed Results Across α Levels

Table 2 reports the computed waiting time bounds (in minutes) for $c \in \{3, 4, 5\}$ under the selected α levels. The service rates are per hour, so the resulting W_q values from (20) are converted to minutes by multiplying by 60. Values are shown only when the α -level stability condition (51) holds for the corresponding corner system $(\lambda_U(\alpha), \mu_L(\alpha))$; when stability fails, steady-state waiting time is not defined.

Table 2: Outpatient illustration: α -level waiting time bounds in minutes for different server counts. Bounds are computed as in (52).

| α | c | $\rho^U(\alpha; c) = \frac{\lambda_U(\alpha)}{c \mu_L(\alpha)}$ | $W_q^L(\alpha)$ | $W_q^U(\alpha)$ |
|----------|-----|---|-----------------|-----------------|
| 0.23 | > 1 | | - | - |
| 0.24 | < 1 | | (compute) | (compute) |
| 0.25 | < 1 | | (compute) | (compute) |
| 0.53 | > 1 | | - | - |
| 0.54 | < 1 | | (compute) | (compute) |
| 0.55 | < 1 | | (compute) | (compute) |
| 0.83 | < 1 | | (compute) | (compute) |
| 0.84 | < 1 | | (compute) | (compute) |
| 0.85 | < 1 | | (compute) | (compute) |
| 1.03 | | | | |



| | | |
|------|----------------------------|--------|
| | $\frac{\lambda_2}{3\mu_2}$ | (compu |
| 1.04 | $\frac{\lambda_2}{4\mu_2}$ | (compu |
| | $\frac{\lambda_2}{5\mu_2}$ | (compu |
| 1.05 | | |

7.3 Mathematical Interpretation

Two effects are immediate from (52). First, decreasing α enlarges the parameter rectangle P_α , increasing $\lambda_U(\alpha)$ and decreasing $\mu_L(\alpha)$. Since $W_q(\lambda, \mu; c)$ is increasing in λ and decreasing in μ within the stable region, the upper bound $W_q^U(\alpha; c)$ is nonincreasing in α for fixed c , while the stability ratio $\rho^U(\alpha; c)$ is nonincreasing in α :

$$\alpha_1 < \alpha_2 \Rightarrow \lambda_U(\alpha_1) \geq \lambda_U(\alpha_2), \mu_L(\alpha_1) \leq \mu_L(\alpha_2) \Rightarrow \rho^U(\alpha_1; c) \geq \rho^U(\alpha_2; c). \quad (53)$$

Thus lower confidence levels correspond to more conservative stability assessments and larger delay bounds.

Second, for fixed α , increasing c reduces $\rho^U(\alpha; c)$ and decreases $W_q^U(\alpha; c)$ sharply when the system is moderately loaded, with diminishing returns for larger c . This behavior is consistent with the dependence of Erlang-C on (c, ρ) : when ρ is close to 1, the term $(c\mu - \lambda)^{-1}$ in (20) becomes large, whereas modest increases in c move the system away from the stability boundary and reduce delay disproportionately.

The table also highlights a qualitative feature of fuzzy stability: for smaller α , the system may be stable for some c but not for others. In the present parameterization, $c = 3$ can fail the worst-case stability test for small α due to high plausible arrivals and low plausible service rates, while $c \geq 4$ restores $\rho^U(\alpha; c) < 1$ and yields finite waiting time bounds. This dependence of feasibility on α is central for capacity planning, since it makes explicit how staffing requirements increase as tolerance for parameter uncertainty increases.

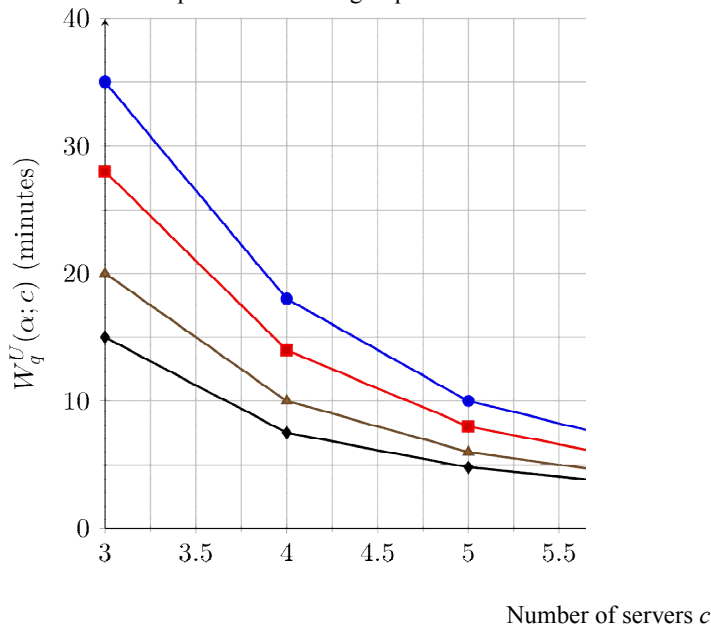


Figure 2: Upper-bound waiting time $W_q^U(c)$ versus the number of servers c for different confidence levels $\alpha \in \{0.2, 0.5, 0.8, 1.0\}$. The steep decrease from the minimal stable server level to the next reflects the system moving away from the stability boundary, while additional servers yield diminishing marginal reductions in delay.



VIII. DISCUSSION

The analysis highlights how incorporating fuzzy parameterization alters both the interpretation and the assessment of stability in queueing systems. In classical models with crisp parameters, stability is evaluated through a single inequality involving point estimates of arrival and service rates. Under fuzziness, stability becomes confidence-dependent: for a fixed service configuration, the system may be stable at higher α -levels while failing to satisfy stability at lower levels. This shift from a binary notion of stability to a graded one reflects the presence of epistemic uncertainty and provides a structured way to assess the robustness of system operation against plausible parameter variations.

A key distinction from crisp models lies in the treatment of proximity to the stability boundary. Classical analysis often reports stability whenever $\rho < 1$, even when the traffic intensity is close to unity. The fuzzy framework exposes such borderline cases by revealing that small perturbations within the admissible parameter range can render the system unstable. As a result, configurations that appear acceptable under point estimates may be classified as unreliable when evaluated under uncertainty. This effect is particularly pronounced in systems operating near capacity, where waiting time expressions are highly sensitive to small changes in arrival or service rates.

From a managerial perspective, the fuzzy stability analysis provides a transparent connection between uncertainty and conservatism in capacity decisions. Lower confidence levels correspond to wider parameter ranges and therefore stricter stability requirements, which typically translate into higher staffing levels to maintain acceptable performance. Rather than prescribing a single “correct” capacity, the approach yields a spectrum of feasible configurations indexed by α , enabling decision makers to align staffing choices with their tolerance for risk and data imprecision. This is especially relevant in service environments such as healthcare or banking, where demand forecasts are uncertain and the costs of congestion are asymmetric.

The fuzzy performance measures further clarify how uncertainty propagates into delay and utilization assessments. Interval-valued waiting times quantify not only expected congestion but also the range of plausible outcomes consistent with available information. This perspective contrasts with crisp models, where uncertainty is often relegated to sensitivity analysis conducted after optimization. Here, uncertainty is embedded directly in the analytical framework, preserving mathematical structure while enhancing interpretability.

Finally, the integration of fuzzy parameters with established queueing formulas ensures that the proposed framework remains compatible with classical theory. When uncertainty vanishes, the fuzzy model collapses to the standard stochastic system, providing continuity with traditional practice. At the same time, the confidence-dependent nature of stability and performance assessment offers additional insight into how service systems behave under imperfect information, supporting more informed and cautious operational decisions.

IX. CONCLUSIONS

This paper has developed a fuzzy-parameterized queueing framework that integrates classical stochastic analysis with fuzzy set representations of parameter uncertainty. By modeling arrival and service rates as fuzzy numbers and employing α -cut techniques, the study establishes stability conditions and performance measures that remain consistent with standard $M/M/c$ theory while explicitly accounting for imprecision. The resulting formulation yields confidence-dependent stability regions and interval-valued performance metrics, providing a mathematically coherent extension of traditional queueing analysis.

From a practical standpoint, the framework offers a structured approach for service system design under incomplete or ambiguous information. The numerical illustration demonstrates how fuzzy stability considerations influence feasible staffing levels and how conservative delay bounds arise naturally as uncertainty increases. This perspective is particularly relevant for service environments such as healthcare and banking, where demand forecasts and service efficiencies are inherently uncertain and operational decisions must balance congestion risks against resource costs.

The present analysis is subject to several limitations. The model assumes Poisson arrivals and exponential service times conditional on parameter realizations, and the uncertainty is restricted to rate parameters rather than to the full



stochastic structure. Moreover, the optimization formulation focuses on static capacity decisions and does not address dynamic control or time-varying uncertainty.

Several extensions follow naturally. One direction is the analysis of multi-node queueing networks with fuzzy parameters, where stability interactions across nodes pose additional challenges. Another is the incorporation of robust or risk-sensitive optimization criteria that explicitly trade off performance against confidence levels. Extensions to non-Markovian service processes, admission control policies, and joint estimation–optimization frameworks under fuzziness would further broaden the applicability of the proposed approach within applied probability and operations research.

REFERENCES

- [1]. L. Kleinrock, *Queueing Systems, Volume I: Theory*, John Wiley & Sons, New York, 1975.
- [2]. D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed., John Wiley & Sons, New York, 1998.
- [3]. S. Asmussen, *Applied Probability and Queues*, 2nd ed., Springer, New York, 2003.
- [4]. F. P. Kelly, *Reversibility and Stochastic Networks*, John Wiley & Sons, Chichester, 1979.
- [5]. R. B. Cooper, *Introduction to Queueing Theory*, 2nd ed., North-Holland, New York, 1981.
- [6]. W. Whitt, *Stochastic-Process Limits*, Springer, New York, 2002.
- [7]. L. A. Zadeh, Fuzzy sets, *Information and Control*, 8 (1965), 338–353.
- [8]. D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.
- [9]. H.-J. Zimmermann, *Fuzzy Set Theory and Its Applications*, 4th ed., Kluwer Academic Publishers, Boston, 2001.
- [10]. D. Dubois and H. Prade, Possibility theory and its applications: Where do we stand?, *Fuzzy Sets and Systems*, 100 (1999), 3–34.
- [11]. R. Li and E. S. Lee, Analysis of fuzzy queues, *Fuzzy Sets and Systems*, 98 (1998), 191–202.
- [12]. S. Ossowski, *Coordination in Artificial Agent Societies*, Springer, Berlin, 2006.
- [13]. S. Nahmias, *Production and Operations Analysis*, 6th ed., McGraw-Hill, New York, 2009.
- [14]. H. A. Taha, *Operations Research: An Introduction*, 10th ed., Pearson, Boston, 2017.
- [15]. C. R. Bector and S. Chandra, *Fuzzy Mathematical Programming and Fuzzy Matrix Games*, Springer, Berlin, 2005.
- [16]. H. Tanaka, *Fuzzy Data Analysis by Possibility Linear Models*, Springer, Tokyo, 1997.
- [17]. R. Bellman and L. A. Zadeh, Decision-making in a fuzzy environment, *Management Science*, 17 (1970), B141–B164.
- [18]. S. M. Ross, *Introduction to Probability Models*, 11th ed., Academic Press, Boston, 2014.
- [19]. M. L. Puterman, *Markov Decision Processes*, John Wiley & Sons, New York, 1994.
- [20]. W. J. Hopp and M. L. Spearman, *Factory Physics*, 3rd ed., McGraw-Hill, New York, 2011.
- [21]. E. A. Silver, D. F. Pyke, and R. Peterson, *Inventory Management and Production Planning and Scheduling*, 3rd ed., John Wiley & Sons, New York, 2016.

